NTNU
Norwegian University of
Science and Technology

TMA4267 Linear Statistical Models V2014 (11)

The multivariate normal density and ML estimates [4.4]
The conditional distribution [4.5]

Mette Langaas

To be lectured: February 10, 2014
wiki.math.ntnu.no/emner/tma4267/2014v/start/

PART 3

random matrix

$\underline{X}$ $p \times 1$ $\qquad$ $\underline{X}$ $n \times p$

random vector

$$\mu = E(\underline{X})$$

$$\Sigma = \mathrm{Cov}(\underline{X}) = E\left[(X-\mu)(X-\mu)^T\right]$$

Var in book

$$Y = C\underline{X}$$

$$E(Y) = C\mu \,, \quad \mathrm{Cov}(Y) = C\,\Sigma\,C^T$$

Properties of Σ: symmetric & positive definite (E3)

$$\Sigma = P \Lambda P^T \,, \quad \Sigma^{1/2} = P \Lambda^{1/2} P^T, \quad \Sigma^{-1} \text{ exists}$$

e_i cols $\qquad$ $\mathrm{diag}(\lambda_i)$

multivariate normal

(mvN)

$$\underset{p\times 1}{\mathbf{X}} \sim N_p(\mu, \Sigma)$$

- linear combinations of $\mathbf{X}$ is mvN
- subsets of $\mathbf{X}$ is mvN
- zero covariance $\Leftrightarrow$ independence
- conditional distributions are mvN

MGF: $M_X(t) = \exp\left(t^T\mu + \frac{1}{2}t^T\Sigma t\right)$

pdf: $f(x) = (2\pi)^{-p/2} \det(\Sigma)^{-1/2} \exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}$

Distribution of (quad form): $(\mathbf{X}-\mu)^T\Sigma^{-1}(\mathbf{X}-\mu) \sim \chi^2_p \leftarrow$

MLE $\hat{\mu} = \bar{\mathbf{X}} = \frac{1}{n}\sum_{j=1}^{n} \underset{p\times 1}{\mathbf{X}_j}$, $S_n = \frac{1}{n}\sum_{j=1}^{n}(X_j - \bar{X})(X_j - \bar{X})^T$

The Multivariate Normal Density [4.4]

- Univariate normal, $X \sim N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}$$

- Multivariate normal:
  Let $\boldsymbol{X}_{(p \times 1)}$ be a random vector from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{p}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} exp\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\}$$

The mvN density [4,4] (cont.)

$$\underset{1\times 1}{Z} \sim N(0,1)$$

$$\Big\downarrow \text{ indep}$$

$$\underset{p\times 1}{Z} \sim N_p(0, I) \xrightarrow{\quad X = \Sigma^{1/2} Z + \mu \quad} X \sim N_p(\mu, \Sigma)$$

$$f(z) = \prod_{i=1}^{p} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} z_i^2} \xrightarrow{\text{mv trans formula}} f(x) = (2\pi)^{-p/2} \det(\Sigma)^{-1/2} \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right\}$$

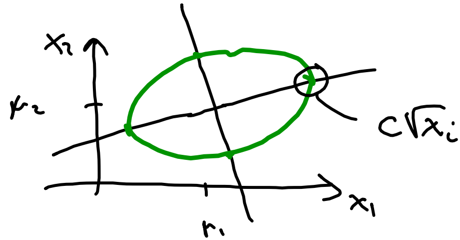
Byproduct (E3)

$$Z = \Sigma^{-1/2}(X-\mu) \sim N_p(0, I) \quad \text{so}$$

$$\underbrace{(X-\mu)^T \Sigma^{-1} (X-\mu)} = Z^T Z = \underbrace{Z_1^2}_{\chi_1^2} + Z_2^2 + \dots + Z_p^2 \sim \chi_p^2$$

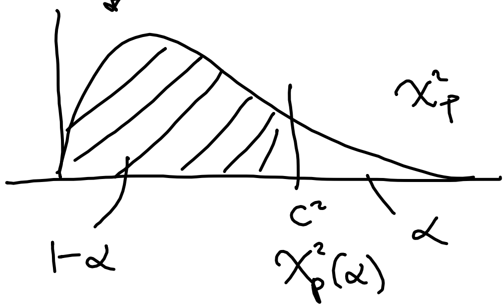
independent

contours of mvN with equal probability are ellipsoids, centered in μ with axes in $e_1, \dots, e_p$ (eigenvectors of Σ).



We also know that

$$P(\underbrace{(X-\mu)^T \Sigma^{-1} (X-\mu) \le c^2}_{\text{points inside ellipoid}})$$



Contours of multivariate normal distribution

- Contours of constant density for the p-dimensional normal distribution are ellipsoids defined by $\boldsymbol{x}$ such that

  $$(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = c^2$$

  These ellipsoids are centered at $\boldsymbol{\mu}$ and have axes $\pm c\sqrt{\lambda_i}\boldsymbol{e}_i$, where $\boldsymbol{\Sigma}\boldsymbol{e}_i = \lambda_i\boldsymbol{e}_i$, for $i = 1, ..., p$.
- $(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$ is distributed as χ^2_p.
- The volume inside the ellipsoid of $\boldsymbol{x}$ values satisfying

  $$(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \leq \chi^2_p(\alpha)$$

  has probability $1-\alpha$.

Maximum likelihood estimators

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be a random sample of size n from the multivariate normal distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The maximum likelihood estimators for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are

$$\bar{\boldsymbol{X}} = \frac{1}{n}\sum_{j=1}^{n}\boldsymbol{X}_j$$

$$\boldsymbol{S}_n = \frac{1}{n}\sum_{j=1}^{n}(\boldsymbol{X}_j - \overline{\boldsymbol{X}})(\boldsymbol{X}_j - \overline{\boldsymbol{X}})^T$$

Maximum likelihood (ML) estimation for mvN

Let $\underset{p\times 1}{X_1}, \dots, X_n$ be a random sample from $N_p(\mu, \Sigma)$.

i.e. $X_1, \dots, X_n$ i.i.d. $N_p(\mu, \Sigma)$

Aim: estimate μ and Σ. ML$\underset{\text{estimator}}{\underset{\uparrow}{\text{E}}}$ maximizes the $L(\mu, \Sigma; x_1, \dots, x_n)$

$$L(\mu, \Sigma; x_1, \dots, x_n) = \prod_{j=1}^{n} f(x_j; \mu, \Sigma)$$

$$= \prod_{j=1}^{n} \left\{ (2\pi)^{-p/2} \det(\Sigma)^{-1/2} \exp\left(-\frac{1}{2}(X_j - \mu)^T \Sigma^{-1} (X_j - \mu)\right) \right\}$$

$$= (2\pi)^{-np/2} \det(\Sigma)^{-\frac{n}{2}} \exp\left(-\frac{1}{2} \underbrace{\sum_{j=1}^{n} (X_j - \mu)^T \Sigma^{-1} (X_j - \mu)}\right)$$

Tricks:

$$\operatorname{tr}(AB) = \operatorname{tr}(BA)$$

$$\operatorname{tr}(A+B) = \operatorname{tr}(A) + \operatorname{tr}(B)$$

$$\operatorname{tr}(\text{scalar}) = \text{scalar}$$

$$\sum_{j=1}^{n} (x_j - \bar{x} + \bar{x} - \mu)(x_j - \bar{x} + \bar{x} - \mu)^T$$

$$= \sum_{j=1}^{n} (x_j - \bar{x})(x_j - \bar{x})^T + n(\bar{x} - \mu)(\bar{x} - \mu)^T$$

homework

$tr(A)$ = sum of the diagonal elements of A

$tr(\Sigma)$ = sum of variances

may see p 113 of BF(2010)

Result of tricks -

$$L(\mu, \Sigma; x_1, \dots, x_n) = (2\pi)^{-\frac{np}{2}} \det(\Sigma)^{-\frac{n}{2}}$$

$$\exp\left\{-\frac{1}{2}\left[tr\left(\Sigma^{-1}\sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})^T\right) + n(\bar{x}-\mu)^T\Sigma^{-1}(\bar{x}-\mu)\right]\right\}$$

<u>Maximum wrt μ</u>

$$\hat{\mu}_{MLE} = \frac{1}{n}\sum_{j=1}^{n} X_j$$

Σ^{-1} is SPD so that $a^T\Sigma^{-1}a > 0 \;\; \forall \; a \neq 0$, so

$a = (\bar{x}-\mu) = 0 \iff \mu = \bar{x}$ gives the max wrt μ.

Maximum wrt Σ

Book: $\frac{\partial L}{\partial \Sigma}$. First insert $\mu = \bar{x}$.

$$\boxed{L \propto \det(\Sigma)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2} \operatorname{tr}\left(\Sigma^{-1} \sum_{j=1}^{n} (x_j - \bar{x})(x_j - \bar{x})^T\right)\right\}}$$

Else use result:

$$\max_{\Sigma} \frac{1}{\det(\Sigma)^{b}} \exp\left(-\operatorname{tr}\left(\Sigma^{-1} B \tfrac{1}{2}\right)\right)$$

is achieved for $\Sigma = \frac{1}{2b} B$ where B is PD (Johnson & Wichern p 170–171)

We have: $b = \frac{n}{2}$, $B = \sum_{j=1}^{n} (x_j - \bar{x})(x_j - \bar{x})$

$$S_n = \hat{\Sigma}_{MLE} = \frac{1}{n} \sum_{j=1}^{n} (X_j - \bar{X})(X_j - \bar{X})^T$$

Properties:

1) $\bar{X} \sim N_p(\mu, \frac{1}{n}\Sigma)$

2) $\dfrac{S_n n}{n-1}$ is an unbiased estimator for Σ.

Properties of the ML estimators

- $\bar{\boldsymbol{X}}$ is distributed as $N_p(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma})$
- $n\boldsymbol{S}$ is distributed as a Wishart random matrix with $n-1$ degrees of freedom.
- $\bar{\boldsymbol{X}}$ and $n\boldsymbol{S}$ are independent.

The Wishart distribution is not on the reading list for TMA4267. General properties of maximum likelihood estimation is covered in detail in TMA4295 Statistical Inference.

Conditioning [4.5]

Ch 1: L3

$$(X,Y) \sim N_2\left(\begin{pmatrix}\mu_x \\ \mu_y\end{pmatrix}, \begin{bmatrix}\sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2\end{bmatrix}\right)$$

The condilion distribution of Y given X=x $\quad f(y|x) = \dfrac{f(x,y)}{f(x)}$

We found $(Y|X=x) \sim N_1(E(Y|X=x), Var(Y|X=x))$

$$E(Y|X=x) = \mu_y + \rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)$$

$$Var(Y|X=x) = (1-\rho^2)\sigma_y^2$$

Which was used to motivate a simple linear regression

$$Y = E(Y|X=x) + \varepsilon$$

$\nwarrow N(0, Var(Y|X=x))$

How to generalize to p dimensions?

$\underset{p\times 1}{\underline{X}} \sim N_p(\mu, \Sigma) \rightarrow$ divide into $\underset{r_1\times 1}{X_1}$ and $\underset{r_2\times 1}{X_2}$

$r_1 + r_2 = p$

Eg. $\underline{X}_1 = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_{r_1} \end{bmatrix}$, $\underline{X}_2 = \begin{bmatrix} X_{r_1+1} \\ \vdots \\ X_p \end{bmatrix}$

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_{r_1} \\ \mu_{r_1+1} \\ \vdots \\ \mu_p \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

$$\Sigma_{11} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1r_1} \\ \vdots & & & \\ \sigma_{r_1 1} & \cdots & - & \sigma_{r_1 r_1} \end{bmatrix} \qquad \Sigma_{22} = \begin{bmatrix} \sigma_{r_1+1, r_1+1} & \cdots & \sigma_{r_1+1, p} \\ \vdots & & \\ \sigma_{p, r_1+1} & & \sigma_{pp} \end{bmatrix}$$

$$\Sigma_{12} = \begin{bmatrix} \sigma_{1, r_1+1} & \cdots & \sigma_{1p} \\ \sigma_{2, r_1+1} & \cdots & \sigma_{2p} \\ \sigma_{r_1, r_1+1} & & \sigma_{r_1, p} \end{bmatrix}, \quad \Sigma_{21} = \Sigma_{12}^T$$

Theorem 4.25

$$\mathbb{X}_2 \mid \mathbb{X}_1 = x \sim N_{r_2}\left(E(\mathbb{X}_2 \mid \mathbb{X}_1 = x_1), \operatorname{Cov}(\mathbb{X}_2 \mid X_1 = x_1) \right)$$

$$E(\mathbb{X}_2 \mid \mathbb{X}_1 = x_1) = \mu_2 + \Sigma_{21} \Sigma_{11}^{-1} (x_1 - \mu_1)$$

$$\operatorname{Cov}(\mathbb{X}_2 \mid X_1 = x_1) = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$$

this motivates multiple linear regression

Proof: p 119 of BF(2010).

kahoot.it

game pin

nick

Properties of the Multivariate Normal Distribution [4.3-4.5]

Let $\boldsymbol{X}_{(p \times 1)}$ be a random vector from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

1. Linear combinations of components of $\boldsymbol{X}$ are (multivariate) normal.
2. All subsets of the components of $\boldsymbol{X}$ are (multivariate) normal.
3. Zero covariance implies that the corresponding components are independently distributed.
4. The conditional distributions of the components are (multivariate) normal.
   $\boldsymbol{X}_2 | (\boldsymbol{X}_1 = \boldsymbol{x}_1) \sim N_{r2}(\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1), \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})$